

OPENCOURSEWARE  
ADVANCED PROGRAMMING  
STATISTICS FOR DATA SCIENCE  
Ricardo Aler



## Stan

Stan is a probabilistic programming language for statistical inference. It is written in C++, so it is quite fast. It follows Bayesian statistics principles. Stan can be used from Python (pystan) or R (**Rstan**). In this course, the R syntax will be introduced. In order to use Stan fully, concepts about Bayesian inference and sampling must be understood. Therefore this lecture is only a small introduction to Stan by explaining its main syntactic issues and a few simple examples.

Standard machine learning or statistical techniques, just fit the model to the data and give some values to the parameters of the model. Stan goes further and returns not just the parameter values, but their distribution. In fact, it returns the full joint probability distribution. This gives some idea about the uncertainty of the parameters (coefficient values in a linear model).

Part of the lecture is devoted to show how Stan achieves this, by using Bayes theorem.

Finally, the practical parts of using RStan are introduced:

- The sections of a basic Stan program
- How Stan can be called from R
- How Stan results can be analyzed, including convergence
- How priors can be added to improve Stan results

**Associated Material :**

- Slides and some exercises (check the lecture notes).